Of-genitive versus s-genitive: A corpus-based analysis of possessive constructions in 20thcentury English

Röthlisberger, Melanie ; Schneider, Gerold

Abstract: This paper examines genitive variation in English, using two methodological approaches. In the manual approach, we extract genitive variants from the parsed sub- corpora of the text category J (academic writing) in the B-Brown (1931), the Brown (1961) and the Frown (1991/2) corpora. Focussing on the syntactic parameter, we illustrate how the principle of end-weight gains ground from 1930 to 1990. The automatic approach implements the constraints of the manual approach, confirms the findings of the manual approach and is used to scale to British English. Methodologically, we show how to automatically sift out irrelevant corpus examples whose identification would normally need human intervention – in particular, apparent examples of the two main genitive English constructions which are not in genuine alternation.

Korpuslinguistik und interdisziplinäre Perspektiven auf Sprache

Corpus Linguistics and Interdisciplinary Perspectives on Language

Bd./Vol. 3

Paul Bennett / Martin Durrell
Silke Scheible / Richard J. Whitt (eds.)

New Methods in Historical Corpora

narr
VERLAG

Bibliografische Information der Deutschen Nationalbibliothek

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über http://dnb.dnb.de abrufbar.

Internet: www.narr.de
E-Mail: info@narr.de

Layout: Andreas Scholz, Essen; www.a-shots.de
Redaktion: Dr. Jacqueline Mullender, Stourbridge UK
Printed in Germany

ISSN 2191-9577
ISBN 978-3-8233-6760-4

MELANIE RÖTHLISBERGER / GEROLD SCHNEIDER

Of-genitive versus *s*-genitive

A corpus-based analysis of possessive constructions in 20th-century English

Abstract

This paper examines genitive variation in English, using two methodological approaches. In the manual approach, we extract genitive variants from the parsed subcorpora of the text category J (academic writing) in the *B-Brown* (1931), the *Brown* (1961) and the *Frown* (1991/2) corpora. Focussing on the syntactic parameter, we illustrate how the principle of end-weight gains ground from 1930 to 1990. The automatic approach implements the constraints of the manual approach, confirms the findings of the manual approach and is used to scale to British English. Methodologically, we show how to automatically sift out irrelevant corpus examples whose identification would normally need human intervention – in particular, apparent examples of the two main genitive English constructions which are not in genuine alternation.

1. Introduction

The increase of *s*-genitives (e.g. *my father's house*) at the expense of *of*-genitives (e.g. *the house of my father*) in modern English is a phenomenon that has received increasing attention. Corpus-based works by Altenberg (1982), Jucker (1993), Rosenbach (2002), and Szmrecsany/Hinrichs (2007) have focussed on the interchangeability between the *s*-genitive and the *of*-genitive in areas where both constructions can be chosen. The choice is constrained by a number of conditioning factors: language internal (i.e. syntactic, lexical, phonological, semantic) as well as external (i.e. factors related to processing, economy-related factors, and socio-stylistic factors). This study contributes to previous work by offering a quantitative analysis of three subcorpora of the *Brown*-family, namely *B-Brown* (1931), *Brown* (1961) and *Frown* (1991/2), while focusing on the syntactic parameter only.

We apply two methodological approaches. In a first step, the genitives are extracted from the syntactically parsed subcorpora, and the data manually edited within the context of interchangeability. Manual filtering is a necessary and time-consuming preliminary to an analysis of factors constraining the

choice. An automated procedure would thus be desirable, particularly when dealing with large corpora. Therefore, in a second step, the fine-grained methods from the first step are approximated by an automatic programming-based approach. The aim of the second step is to incorporate previous manual work into an automatic work flow.

This two-fold methodological approach offers unique insights into the possible applicability of manually-applied constraints to computerised automatic searches, and as a consequence, the possible extension of the scope of research to larger amounts of data, to new genres and new varieties.

2. Previous research

Previous research has so far focused on the various parameters that influence the choice of genitive constructions whereby the set of parameters and their relative importance often differs from scholar to scholar (see Szmrecsanyi/Hinrichs 2007: 438). Jucker (1993) counted six factors based on Altenberg (1982): the phonetic, morphological, syntactic, lexical, and relational factors and the degree of formality. Szmrecsanyi/Hinrichs (2007) followed Jucker (1993) and analyzed their data under four major conditioning factors, taking the syntactic and pragmatic levels, as well as communicative aspects and language processing, into account. Rosenbach's (2003) influential work on genitive choice takes only three factors into account: animacy, topicality and possessive relation. She categorically excludes any factors that bias the free choice between the two genitive variants.

The availability of a set of corpora stretching across three time periods and spanning more than half a century offered a unique opportunity to apply the methodological approaches of previous research to a new dataset, and hence give insights into the diachronic changes in genitive choice in American English from 1930 to 1960 to 1990. A previous pilot study with this dataset (Röthlisberger 2009) led to the conclusion that the syntactic factor is one of the most influential parameters in genitive choice, and this has hence been chosen as the focus of this study.

3. Manual Approach

3.1 Data

We used the three corpora *B-Brown*, *Brown* and *Frown* in order to analyse the changes in frequency between 1930 and 1990, and to detect possible differences in change between the periods 1930-1960 and 1960-1990. Within these three corpora we focused on category J (Academic writing). We had to rely on those subcorpora of American English, because the B-Brown corpus is still not completed, and the J and K texts are the only ones fully available in their parsed version.

Surface pattern searches incur a large error rate, since both *s*-genitive forms and *of*-prepositions are ambiguous. The surface form of *s*-genitives is identical to contracted forms of *be* in the 3rd person singular (e.g. *Peter's painting is large* vs. *Peter's painting a house*), and the attachment site of prepositions is ambiguous. In *We accused the man of robbery* the *of*-preposition attaches to the verb, in *the state of emergency of the nation* the second *of*-preposition attaches to *state* and not to *emergency*. In order to minimise the error rate, each text was parsed with the syntactic parser *Pro3Gres* (Schneider 2008). Figure 1 shows an example of the parser output. The *s*-genitive *a layperson's point* is annotated with the dependency label *pos* (possessive), the *of*-genitive point of view with the dependency label *modpp* (modification by PP; this label is used for noun-PP attachment). The possessum is the governor in the relation, the possessor the dependent. Parsing approaches are not error-free either (see Schneider 2008). For the *s*-genitive and *of*-genitive, the accuracy on the 500 sentences *GREVAL* evaluation corpus (Carroll et al. 2003) is as follows: *of*-PP precision=89.6% and recall=85.82%. *s*-genitive precision and recall=97.4%.

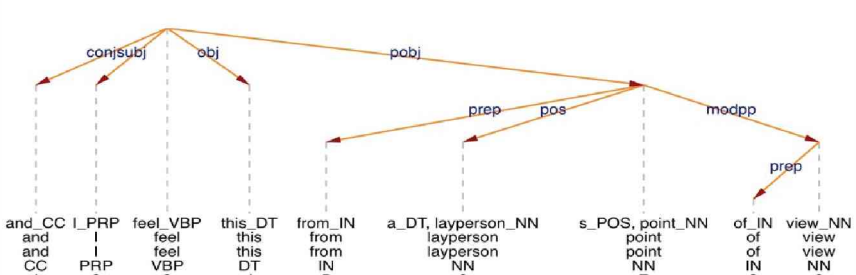

Figure 1: Parser output of a sample sentence.

3.2 Methodology

To analyse the variation between *of*- and *s*-genitive within the three subcorpora, the whole set of genitives was restricted to constructions occurring in so-called 'choice context', which goes back to Labov's principle of accountability of (1969). Rosenbach (2002) defines "choice context" as the linguistic context where both genitive variants can occur in free variation (Rosenbach 2002:40). To extract the relevant interchangeable genitive constructions, this study relies heavily on the methods used by Rosenbach (2002: 28ff), Szmrecsanyi (2007:448), Ljung (1997: 30), Raab-Fischer (1995: 127), and Kreyer (2003: 170-171).

A conversion rule was applied to all genitives found in the dataset: both the original and its alternative genitive construction have to be semantically equivalent and grammatically correct with both possessum and possessor as nouns (thus, for example, excluding genitives with pronominal possessors). The genitives need to have a possessive genitive function and should not appear in an idiomatic expression or conventionalised phrase. The excluded constructions are descriptive genitives, independent genitives, local genitives, post-genitives, nested and group genitives, elliptic genitives, *s*-genitive construction whose possessum is premodified by *own*, and titles of books, films, and works of art that are premodified by their creator's name. In order to bring forth a comparable context, *of*-genitives with a referential device other than a definite element are excluded from the analysis (e.g. *a nest of a bird*) because *s*-genitives are already definite in their nature (Langacker 1995:63). We also exclude almost all *of*-genitives with a possessor that shows a clausal postmodification, due to the conversion rule. Additionally, we exclude measures expressed with *of*-constructions, and *of*-constructions where the possessum modifies the possessor, because such constructions result in an ungrammatical descriptive genitive when converted (e.g. *a king of honour* ≠ *an honour's king*). The data was manually filtered by applying the conversion rule and the set of restrictions as noted above. The three parsed subcorpora were analysed with *SWI-Prolog*, where we programmed a rules file according to the constraints of this study.

3.3 Syntactic factors: theoretical approach

On the level of each constituent, both possessor and possessum can constitute a noun phrase with its governor and modifications (Kreyer 2003: 179). The principle of end-weight and the proximity principle (we do not discuss the lat-

ter here) are two of the most important syntactic factors that affect the individual distribution of pre- and postmodifications on the level of the possessive noun phrase, and on the level of possessor and possessum. This principle implies that longer phrasal constituents tend to follow shorter ones (Szmrecsanyi/Hinrichs 2007: 453). Therefore, a pre-or postmodified possessor-NP should favour *of*-genitive as in 1), whereas a pre-or postmodified possessum-NP favours *s*-genitive as in 2).

1) *the centre* [possessum] *of a guarded heart* [possessor]

2) *my mind's* [possessor] *ability to communicate* [possessum]

3.4 Results

The application of the aforementioned constraints produced the following frequencies of *s*- and *of*-genitives in the three subcorpora:

	s-genitives		*of*-genitives		TOTAL
	N	%	N	%	
B-Brown-J	162	11.0	1306	89.0	1468
Brown-J	179	18.0	814	82.0	993
Frown-J	352	34.4	670	65.6	1022

Table 1: *s*- and *of*-genitives in academic writing in *B-Brown*, *Brown* and *Frown*

	1930-1960	1960-1990
s-genitives	+10.5%	+96.7%
of-genitives	−37.7%	−17.7%

Table 2: Changes in frequency from 1930-1960 and 1960-1990 by genitive type

Table 2 suggests that *of*-genitives decreased to a greater extent in the time period 1930-1960, thus possibly creating a functional and syntactic gap that is filled by the *s*-genitive in the period 1960 to 1990. However, data is too sparse to be able to make a specific claim for such a drag-chain. The differences in frequencies in *B-Brown*, *Brown*, and *Frown* (Table 1) are very highly significant ($df=2$, $p<0.0001$, $\chi^2=2154.06$).

In order to assess the influence of weight and the principle of end-weight, we measure important features related to weight in the following, in particular post- and premodification (Jucker 1993), length and relative length (Szmrecsanyi/Hinrichs 2007).

3.4.1 Syntactic factors: pre- and postmodification

In a first step, the genitive constituents were filtered according to their pre- or postmodifications. Any referential device modifying the possessor in an *of*-construction did not count as modification (e.g. *the motivations of **the** actors*) (cf. Szmrecsanyi/Hinrichs 2007: 453). For methodological reasons, a compound consisting of Noun+Noun was considered to have a modification – namely the first noun. We analysed all genitives according to the pre- and/or postmodifications of their constituents (Table 3).

Type of modification	*s*-gen			*of*-gen		
	1930	**1960**	**1990**	**1930**	**1960**	**1990**
No modification	124	97	204	579	279	227
Premodification of possessum	29	53	119	300	185	132
Premodification of possessor	6	17	18	279	190	198
Postmodification of possessum	0	2	0	0	1	0
Postmodification of possessor	0	0	0	0	2	1
Post-and Premod. of possessum	0	0	0	0	0	0
Post- and premod. of possessor	0	0	0	0	0	2
Modification of possessum and possessor	3	10	11	148	157	110
TOTAL	**162**	**179**	**352**	**1306**	**814**	**670**

Table 3: Distribution of modified possessors and possessums by corpus and genitive type

When applying the χ^2-test, the difference between the frequencies of *s*-genitives from 1930 to 1990 is highly significant (df=6, χ^2=31.44, p<0.001). For the *of*-genitive, the χ^2-test at df=6 gives χ^2=56.37, and p=0; the differences from 1930 to 1990 in the choice of *of*-genitives are also highly significant. For a further analysis, we only took those lines into account in which more than half of the numbers are higher than 5, and compared the differences in the corpora between 1930-1960 and 1960-1990, using the log-likelihood test (Table 4).[1]

[1] The critical values to indicate significance are: $p < 0.05$; critical value = 3.84 *; $p < 0.01$; critical value = 6.63 **; $p < 0.001$; critical value = 10.83 ***; $p < 0.0001$; critical value = 15.13 ****

	s-genitives		*of*-genitives	
	1930-1960	1960-1990	1930-1960	1960-1990
No modification	−6.55*	+0.30	−12.82***	−0.02
Premodification of possessum	+4.94*	+0.66	−0.01	−1.58
Premodification of possessor	+4.44*	−3.28	+0.88	+5.39*
Mod. of possessum and possessor	+3.31	−1.73	+21.41****	−1.69

Table 4: The log-likelihood value between the different corpora according to modifications

The significant increase of *s*-genitives with premodified possessums and *of*-genitives with premodified possessors indicates that the principle of end-weight gains ground. Note that this increase occurs for *s*-genitives in the period 1930-1960, and *of*-genitives in the later period. This change could therefore be interpreted as a push-change that starts with the *s*-genitive. The significant increase of *s*-genitives with premodified possessors in the period 1930-1960 runs counter to the concept of end-weight. Further research will be needed in that direction.

3.4.2 Syntactic factors: constituent length

In a second step, we established the boundaries of each genitive construction and calculated the mean possessor and possessum length in orthographic words. Any referential device modifying the possessum or possessor was again not taken into account (Table 5).

Corpora	*s*-genitive		*of*-genitive	
	Mean N1 length	Mean N2 length	Mean N1 length	Mean N2 length
B-Brown-J	1.06	1.23	1.48	1.44
Brown-J	1.18	1.44	1.57	1.52
Frown-J	1.09	1.47	1.65	1.44

Table 5: Mean possessor (N1) and possessum (N2) length in the three corpora

Table 5 illustrates that the mean length of the first constituent in *s*-genitives and *of*-genitives remains fairly stable across the years, while the last constituent in both constructions tends to increase in length. Note that the last constituent in *s*-genitive is N2, while in *of*-genitive it is N1. Again, this points to the influence of end-weight.

A comparison between the lengths of possessor and possessum demonstrates that a difference in length influences the choice of genitive (Table 6).

	s-genitives		*of*-genitives	
	1930-1960	1960-1990	1930-1960	1960-1990
N1>N2	+5.14*	−4.04*	+0.19	+6.22*
N1=N2	−4.30*	+0.02	−1.09	−1.01
N1<N2	+5.18*	+0.56	+1.06	−2.10

Table 6: Log-likelihood test for length of possessor/possessum according to time period

Table 6 indicates that *s*-genitives with a longer first constituent (N1>N2) increase significantly from 1930 to 1960, while the same holds true for *s*-genitives with a longer last constituent (N1<N2). Only the later change follows the principle of end-weight. The first change is either caused by other factors, or may be due to low counts (N=6 is lowest for N1>N2 in 1930, while N=30 is lowest for N1<N2 in 1930). From 1960 to 1990, *s*-genitives with longer first constituents decrease significantly, this time following the principle of end-weight. The principle also seems to hold true for the increase of *of*-genitives with longer last constituents in the time period 1960-1990. The changes across the whole table are highly significant (χ^2 contingency table, df=10, $p < 0.001$).

Overall, the manual approach has shown that the changes in genitive choice tend to follow the principle of end-weight and are generally significant.

4. Automatic approach

We have already automated the syntactic annotation in the manual approach, which constitutes a new method in historical corpora. In this section, we suggest automatic approaches approximating to the manual approach described in Section 3. Automatic approaches have the advantage that they scale, and are consistent and reproducible.

4.1 Methods

We discussed our parsing method in Section 3.1. Only a subset of the Saxon genitives (*s*-genitives) and *of*-PPs are in variation. As the envelope of variation (Labov 1969), which Rosenbach (2003) calls the choice context, is subject to semantic restrictions, its automation is challenging. We now suggest approximations and discuss results in Section 4.2.

4.1.1 Raw counts

Assuming that occurrences of variation and non-variation of the *s*-genitive and *of*-genitive are spread homogeneously across the corpus, raw counts can be used as a coarse measure.

4.1.2 Animacy and proper names

In prototypical *s*-genitives, the possessor is a proper name. Restricting counts to cases where the possessor is a proper name is thus a useful approximation: a large portion of the cases in the variation are covered, and only a few false positives included. Proper names and animacy are related.

4.1.3 Data-driven alternations

The only reliable proof of variation is to test if a token can be in the alternation, in other words that both the original and its alternative genitive construction have the same meaning, which we have tested in the manual approach. As this test relies on semantics and speaker intuition, it cannot be automated easily.

Idioms	*point of view* <≠> **view's point* **(eye) view of bird* <≠> *bird's (eye) view*
Creators	*Spielberg's film* <≠> *?film of Spielberg*
Fixed nominal expressions / Proper names	*Noah's ark* <=> *?Ark of Noah* *Newton's comet* <=> *?Comet of Newton* *Institute of Archaeology* <=> **Archaeology's Institute*
Measures / Quality	*tin of soup* <≠> **soup's tin* *half of (the) century* <≠> **century's half*
Semantic restrictions	*one's recovery* <=> **recovery of one* *God's creation* <≠> *?creation of God*
... many other expressions that are not in the alternation, e.g.:	*image of power* <≠> *?power's image* *concentration of oxygen* <=> *?oxygen's concentration* *faculty of reason* <≠> *?reason's faculty*

Table 7: Examples of automatically excluded alternation candidates

What we can test, however, is whether the alternative form does occur in the corpus. If the two alternatives with the same lexemes are found in the corpus, they constitute a valid alternation.

(LEX) B's A <=> A of B

For example, if both the NP *Peter's friend* and *friend of Peter* are found in the corpus, then they are a valid alternation. There are two differences between such an automatic test and the manual approach. First, the automatic approach is based on performance instead of competence. Second, the requirement of semantic equality cannot be tested, so some false positives will be generated.

4.1.4 Adding semantic classes to overcome sparse data

In practice, there is a third difference. There is typically a serious sparse data problem in that there are relatively few pairs with lexical overlap of both the possessum (governor) and the possessor (dependent). In order to alleviate this problem, we require semantic class overlap instead of lexical overlap.

(SEM.1) B's A <=> class(A) of class(B)

For example, if both the NPs *Peter's friend* and *wife of John* are found in the corpus, they are accepted as valid alternation, because *Peter* and *John* are in the same semantic class, as well as *wife* and *friend*. As lexical class, we use the *WordNet* lexicographer file (Miller 1990). Semantic class overlap shows high correspondence with manual decisions.[2] Classes of automatically excluded pairs are given in Table 7.

One of the restrictions of the manual approach can be automated directly: *of*-genitives with a referential device other than a definite element need to be excluded from the analysis (e.g. *a nest of a bird*). We have added this restriction.

(SEM) B's A <=> class(A) of class(B) AND B is definite

4.2 Results

4.2.1 Raw counts

The raw counts (RAW) are compared in Table 8. The *s*-genitive increases, and *of*-PP seems to decrease, which is in accord with Leech et al. (2009:48) and with our manual data (see Table 2). The increases and decreases are also shown in Table 8. Column 2 gives the ratio (*s*- divided by *of*-), columns 3 and 5 absolute counts, and columns 4 and 6 give percentages.

Corpus	*s*/*of*	*s*-gen #	%	*of*-gen #	%
RAW					
B-Brown-J	0.05	347	4.7%	6998	95.3%
Brown-J	0.06	411	6.1%	6356	93.9%
Frown-J	0.12	716	10.9%	5853	89.1%
1930-60		+18.4%		−9.7%	
1960-90		+74.2%		−7.9%	

Table 8: Raw counts and frequency changes

[2] We did not conduct a formal evaluation of the overlap. Figure 2 gives an indication of the quality.

4.2.2 Animacy and proper names

The proper name counts (PROP) are given in Table 9. Columns 2 and 4 give absolute counts. Columns 3 and 5 give percentages for proper names, showing that proper name genitives are increasingly often realized as *s*-genitives. The readiness for proper name in *of*-PPs is generally low and decreases, as the last two columns show. They give the percentages of genitives with proper names.

Corpus	*s*-gen #	%*s*/ Prop N	*of* #	%*of*/ Prop N	%Prop N/*s*-gen	%Prop N/ *of*-gen
PROP						
B-Brown-J	255	28.7%	634	71.3%	73.5%	9.1%
Brown-J	245	35.4%	447	64.6%	59.6%	7.0%
Frown-J	468	53.7%	404	46.3%	65.4%	6.9%

Table 9: Proper Names counts and percentages

4.2.3 Data-driven alternations

As discussed in Section 4.1.3, the data-driven alternation counts are too low to be reliable or statistically significant, with only between 16 and 35 counts per cell.

4.2.4 Adding semantic classes to overcome sparse data

The counts for semantic class overlap, method SEM.1, are given in Table 10, which can be compared to the results of the manual method given in Table 1. The *of*-genitive overgenerates considerably compared to the manual method.

Corpus	*s*-gen #	% *s*/all	*of* #	% *of*/all
SEM.1				
B-Brown-J	240	6.6%	3 404	93.4%
Brown-J	319	9.5%	3 039	90.5%
Frown-J	564	15.4%	3 101	84.6%

Table 10: Semantic class overlap counts and percentages

The semantic class counts with added indefinite filter, method SEM, are given in Table 11. The *of*-genitive overgenerates less than in SEM.1. The suggested trends are in full agreement with those found by the manual approach.

Corpus	*s*-gen #	% *s*/all	*of* #	% *of*/all
SEM				
B-Brown-J	240	9.8%	2 207	90.2%
Brown-J	319	15.5%	1 737	84.5%
Frown-J	564	25.6%	1 637	74.4%
1930-60		32.9%		−21.3%
1960-90		76.8%		−5.8%

Table 11: Semantic class overlap counts and percentages

4.3 Scaling to British English and other genres

An advantage of automatic approaches is that they scale to other genres, and different, and larger, corpora. We extended our investigation to category K, and to the *LOB* series of corpora (*BLOB* (1931), *LOB* (1961) and *Freiburg LOB* (1991)). We show the results using raw counts (RAW) in Table 12, and semantic class plus indefinite filter (SEM) in Table 13.

Corpus	*s*-gen #	% *s*/all	*of* #	% *of*/all
RAW				
BLOB-J	362	5.0%	6 929	95.0%
LOB-J	425	6.7%	5 897	93.3%
FLOB-J	575	9.0%	5 824	91.0%

Table 12: Raw counts in the *LOB* family

Corpus	*s*-gen #	% *S*/all	*of* #	% *of*/all
SEM				
BLOB-J	243	12.5%	1 706	87.5%
LOB-J	333	15.8%	1 776	84.2%
FLOB-J	372	18.1%	1 687	81.9%
1930-60		37.0%		4.1%
1960-90		11.7%		−5.0%

Table 13: Changes in frequency in the *LOB* family

The trend is similar for the *LOB* family: *s*-genitive increases, and *of*-genitive decreases, although less strongly than in the Brown family, and only relative to the frequency of the *s*-genitive.

4.4 The principle of end-weight

We also investigated the principle of end-weight using the automatic approach SEM.1. We measured the probability of the tokens being pre-modified, which is comparable to the constituent length in Section 3.4.2, Table 5. The results for *Brown* and *LOB* are given in Table 14. Similar results were obtained: the constituent at the end (in bold) has a much higher likelihood of being modified in the more recent corpora, indicating the increased importance of the principle of end-weight.

Corpus	*s*-gen			*of*-gen		
	N	p(DepMod)	**p(GovMod)**	N	**p(DepMod)**	p(GovMod)
B-Brown-J	240	11.7%	**27.1%**	3404	**43.8%**	33.5%
Brown-J	319	13.5%	**32.9%**	3039	**46.0%**	38.4%
Frown-J	563	11.4%	**35.2%**	3101	**47.7%**	37.0%
BLOB-J	243	18.5%	**16.9%**	2407	**39.0%**	31.6%
LOB-J	333	18.0%	**22.8%**	2700	**44.3%**	34.2%
FLOB-J	372	13.4%	**25.0%**	3023	**47.0%**	33.7%

Table 14: End-weight in the *Brown* and *LOB* families

The differences in *Brown* are highly significant (χ^2 contingency, df=2, $p < 0.001$). The differences in *LOB* are significant, but not highly significant (χ^2 contingency, df=2, p=0.039).

5. Discussion

We investigated changes in *s*-Genitives and *of*-Genitives and the principle of end-weight from different perspectives, and observed the same trends in all perspectives. Comparing the absolute counts delivered by the different methods on the *Brown* series in Table 14, it is clear that raw counts overgenerate massively, while the semantic class + indefinite filter counts overgenerate less, as the raw numbers show. Counts for the manual method (MAN) are listed in Table 2, for the raw count method (RAW) in Table 10, and for the semantic class filter (SEM) in Table 11. The percentage increase of *s*-genitives is compared in Figure 2. The comparison shows that SEM is a better approximation to MAN than RAW.

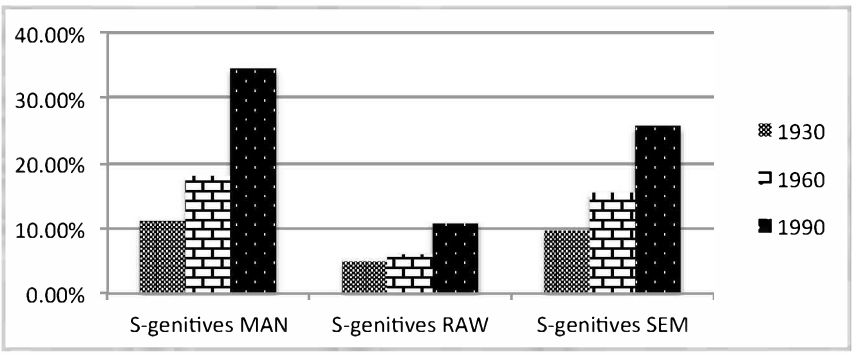

Figure 2: Percentage of Saxon Genitive measured by the different approaches

6. Conclusions

We have shown that the use of the *s*-genitive increased between 1930 and 1990 in both American and British English, while the *of*-Genitive has decreased. The *s*-genitive becomes more restricted to proper names. The differences over the time periods are significant. We have shown that the principle of end-weight has become stronger. We have also presented an approach to the automatic detection of pairs in genitive alternation, which can partly alleviate the workload of the annotator. The manual and automatic approaches are mutually validating. Although automatic approximation overgenerates and delivers a weaker signal, it clearly shows the same trends as the manual approach. In future research, we will conduct a formal evaluation and port the technique to other choice contexts, for example the dative shift.

7. Acknowledgements

We would like to thank two anonymous reviewers for their helpful comments.

References

Altenberg, Bengt (1982): The genitive v. the *of*-construction: a study of syntactic variation in 17th century English. Lund: CWK Gleerup.

Barber, Charles (1964): Linguistic change in Present-Day English. Edinburgh/ London: Oliver & Boyd.

Biber, Douglas (1987): A textual comparison of British and American writing. In: American Speech 62: 99-119.

Carroll, John/Minnen, Guido/Briscoe, Edward (2003): Parser evaluation: using a grammatical relation annotation scheme. In: Abeillé, Anne (ed.): Treebanks: building and using parsed corpora. Kluwer: Dordrecht, 299-316.

Francis, W. Nelson/Kučera, Henry (1982): Frequency analysis of English usage: lexicon and grammar. Boston: Houghton Mifflin Company.

Hawkins, Roger (1981): Towards an account of the possessive constructions. NP's N and the N of NP. In: Journal of Linguistics 17: 247-269.

Jucker, Andreas H. (1993): The genitive versus the *of*-construction in newspaper language. In: Jucker (ed.), 121-136.

Jucker, Andreas H. (ed.) (1993): The Noun phrase in English: its structure and variability. Heidelberg: Winter.

Kreyer, Rolf (2003): Genitive and *of*-construction in modern written English: processability and human involvement. In: International Journal of Corpus Linguistics 8: 169-207.

Langacker, Ronald W. (1995): Possession and possessive construction. In: Taylor, John R./MacLaury, Robert E. (eds.): Language and the cognitive construal of the world. Berlin/New York: Mouton de Gruyter, 51-79.

Labov, William (1969): Contraction, deletion, and inherent variability of the English copula. In: Language 45: 715-762.

Leech, Geoffrey/Smith, Nicolas (2006): Recent grammatical change in written English 1961-1992. Some preliminary findings of a comparison of American with British English. In: Renouf, Antoinette/Kehoe, Andrew (eds.): The changing face of corpus linguistics. Amsterdam/New York: Rodopi, 185-204.

Leech, Geoffrey/Hundt, Marianne/Mair, Christian/Smith, Nicholas (2009): Change in contemporary English: a grammatical study. Cambridge: Cambridge University Press.

Ljung, Magnus (1997): The *s*-genitive and the *of*-construction in different types of English texts. In: Fries, Udo/Müller, Viviane/Schneider, Peter (eds.): From Aelfric to the New York Times. Studies in English Corpus Linguistics. Amsterdam: Rodopi, 21-32.

Mair, Christian (2006): Inflected genitives are spreading in present-day English, but not necessarily to inanimate nouns. In: Mair, Christian/Heuberger, Reinhard (eds.): Corpora and the History of English. Heidelberg: Winter, 235-248.

Miller, George A./Beckwith, Richard/Fellbaum, Christiane/Gross, Derek/Miller, Katherine (1990): Wordnet: an on-line lexical database. In: International Journal of Lexicography 3: 235-244.

Ortmann, Albert (1998): The role of [+/– animate] in inflection. In: Fabri, Ray/Ortmann, Albert/Parodi, Teresa (eds.): Models of inflection. Tübingen: Niemeyer, 60-84.

Osselton, Noel. E. (1988): Thematic genitives. In: Nixon, Graham/Honey, John (eds.): An historic tongue: studies in English linguistics in memory of Barbara Strang. London/New York: Routledge, 138-144.

Raab-Fischer, Roswitha (1995): Löst der Genitiv die *of*-phrase ab? Eine korpusgestützte Studie zum Sprachwandel im heutigen Englisch. In: Zeitschrift für Anglistik und Amerikanistik 43: 123-132.

Röthlisberger, Melanie (2009): *Of*-genitive versus *s*-genitive. A corpus-based analysis of possessive constructions in 20th century American English. Unpublished seminar paper: Universität Zürich.

Rosenbach, Anette (2002): Genitive variation in English. Conceptual factors in synchronic and diachronic studies. Berlin: Mouton de Gruyter.

Rosenbach, Anette (2003): Aspects of iconicity and economy in the choice between the *s*-genitive and the *of*-genitive in English. In: Rohdenburg, Günter/Mondorf, Britta (eds.): Determinants of grammatical variation in English. Berlin/New York: Mouton de Gruyter, 379-411.

Rosenbach, Anette (2005): Animacy versus weight as determinants of grammatical variation in English. In: Language 81: 613-644.

Rosenbach, Anette (2006): On the track of Noun+Noun constructions in Modern English. In: Houswitschka, Christoph/Knappe, Gabriele/Müller, Anja (eds.): Proceedings of the Conference of the German Association of University Teachers of English 27. Trier: Wissenschaftlicher Verlag, 543-557.

Rosenbach, Anette (2007). Exploring constructions on the web: a case study. In: Biewer, Carolin/Hundt, Marianne/Nesselhauf, Nadja (eds.): Corpus linguistics and the Web. Amsterdam/New York: Rodopi, 167-190.

Rosenbach, Anette/Stein, Dieter/Vezzosi, Letizia (2000): On the history of the *s*-genitive. In: Bermudez-Otero, Ricardo/Denison, David/Hogg, Richard M./McCully, Christopher B. (eds.): Generative theory and corpus studies: A dialogue from ICEHL 10. Berlin/New York: Mouton de Gruyter, 185-210.

Schneider, Gerold (2008): Hybrid long-distance functional dependency parsing. Unpublished Dr. Phil. thesis: Philosophische Fakultät, Universität Zürich.

Szmrecsanyi, Benedikt/Hinrichs, Lars (2007): Recent changes in the function and frequency of Standard English genitive constructions: a multivariate analysis of tagged corpora. In: English Language and Linguistics 11: 437-474.

Szmrecsanyi, Benedikt/Hinrichs, Lars (2008): Probabilistic determinants of genitive variation in spoken and written English. A multivariate comparison across time, space, and genres. In: Nevalainen, Terttu/Taavitsainen, Irma/Pahta, Päivi/Kor-

honen, Minna (eds.): The Dynamics of linguistic variation. Corpus evidence on English past and present. Amsterdam/Philadelphia: John Benjamins, 291-309.

The American heritage dictionary of the English language (2000). Fourth edition. Boston: Houghton Mifflin.

Varantola, Krista (1993): Modification of nouns by nouns – bad by definition. In: Jucker (ed.), 69-83.

Wasow, Thomas (1997): Remarks on grammatical weight. In: Language Variation and Change 9: 81-105.